AMENDMENTS TO THE SPECIFICATION:

On page 9 of the BRIEF DESCRIPTION OF THE DRAWINGS amend the description of Figure 6 as follows:

~~Figure 6~~ Figure 6A is a topographic map produced using *VxInsight* showing 9 novel biologic clusters of ALL (2 distinct T ALL clusters (S1 and S2) and 7 distinct B precursor ALL clusters (A, B, C, X, Y, Z)) each with distinguishing gene expression profiles. Figure 6B is a detailed examination of the cluster data of Figure 6A.

On page 9 of the BRIEF DESCRIPTION OF THE DRAWINGS amend the description of Figure 7 as follows:

Figure 7 shows a gene list comparison. Principal Component Analysis ~~(PCA~~ (PCA) and the *VxInsight* clustering program (ANOVA) were employed to identify genes that determined T-cell leukemia cases. The gene lists are compared with those derived from the different feature selection methods used by Yeoh et al. (Cancer Cell, 1:133-143, 2002) for T-cell classification. The ~~yellow color~~ light grey shaded genes ~~represents~~ represent overlap between the lists derived by PCA and the T-ALL characterizing gene lists; the ~~cyan~~ dark grey shaded genes ~~represents~~ represent overlap between the ANOVA and the T-ALL characterizing gene lists. The ~~green pattern represents~~ genes that are shared by all the lists are represented in shaded boxes that have a solid black border.

On page 9 of the BRIEF DESCRIPTION OF THE DRAWINGS amend the description of Figure 8 as follows:

Figure 8 shows a gene list comparison. Bayesian Networks were employed to identify genes that determined the gene expression patterns across the different translocations. The gene lists were compared with those derived using chi square analysis by Yeoh et al. (Cancer Cell, 1:133-143, 2002) for ALL classification. The ~~colored~~ grey shaded cells represent overlap between the lists derived by Bayesian nets and the ALL characterizing gene lists from Yeoh et al. (Cancer Cell, 1:133-143, 2002).

On page 9 of the BRIEF DESCRIPTION OF THE DRAWINGS amend the description of Figure 9 as follows:

Figure 9 shows Principal Component Analysis of the infant gene expression data. Principal Component Analysis (PCA) projections are used to compare the ALL/AML partition, the MLL/Non-MLL partition, and the VxInsight partition of the infant gene expression data. The three by three grid of plots in this figure allows this comparison by using the same PCA projections with different ~~colors~~ shading for the different partitions. Each row of the grid shows a different partition and each column shows a different PCA projection. The ALL/AML partition is shown in the first row of the figure using light ~~purple~~ grey shading for ALL and dark ~~purple~~ grey shading for AML. The three plots in this row give two-dimensional projections of the data onto the first three principal components. Since there are three such projections there are three plots (from left to right): PC 1 vs. PC 2, PC 2 vs. PC 3, and PC 1 vs. PC 3. This scheme is repeated for the

remaining two partitions. Specifically, the MLL/Non-MLL partition is shown using ~~orange and dark green~~ light grey shading and dark grey shading in the second row, and the VxInsight partition is shown using ~~red, green, and blue~~ light grey shading, medium grey shading and solid black fill in the last row. This grid enables both visualization of the data (by examining the rows) and comparison of the partitions (by examining the columns).

On page 10 of the BRIEF DESCRIPTION OF THE DRAWINGS amend the description of Figure 10 as follows:

Figure 10 shows results of the graphic directed algorithm applied to the infant dataset. The *VxInsight* program constructs a mountain terrain over the clusters such that the height of each mountain represents the number of elements in the cluster under the mountain. ~~Top left: this~~ Fig. 10A shows a force-directed clustering algorithm that partitions the infant data into three clusters labeled A, B, and C. ~~Top right:~~ Fig. 10B consists of a *VxInsight* terrain map showing the distribution of the leukemia types across the same clusters. ALL cases are shown in white and AML are shown in ~~green~~ light grey shading. ~~Bottom left:~~ Fig. 10C consists of a *VxInsight* terrain map showing the distribution of MLL cases (shown in ~~blue~~ dark grey shading) across the clusters labeled A, B, and C.

On page 10 of the BRIEF DESCRIPTION OF THE DRAWINGS amend the description of Figure 11 as follows:

Figure 11 shows hierarchical clustering of the 126 infant leukemia samples using the "cluster-characterizing" gene sets. The rows represent genes that distinguish between the *VxInsight* clusters from Figure 2 (n=150). Genes were selected by ANOVA as being the 0.1% top discriminating between each one of the clusters and the rest of the cases. Each gene is normalized across all 126 cases and the relative expression is depicted in the heat map by ~~color~~ shades of grey and solid black, as shown in the expression scale ~~in~~ located at the bottom of ~~the figure~~ Fig. 12A. The patient-to-patient distance was computed using Pearson's correlation coefficient in the Genespring program (Silicon Genetics). The columns in the dendrogram represent patients as clustered by their gene expression. The correlation between these three resultant clusters and the VxInsight clusters is higher than 90%.

On page 10 of the BRIEF DESCRIPTION OF THE DRAWINGS amend the description of Figure 12 as follows:

~~Figure 12~~ Figures 12A and 12B ~~shows~~ show gene expression for various hematopoietic stem cell antigens in the infant leukemia data set. Fig. 12A is a gene expression "heat map" of selected HOX genes and hematopoetic stem cell antigens. The columns represent genes, while the rows represent patients organized by their VxInsight cluster membership A, B or C (see Fig. 10). The gene expression signals of 31 genes from the 26 leukemia patients were normalized relative to the median signal for each gene. The ~~color charcaterizes~~ grey shading characterizes the relative ~~expresssion~~ expression from the median. ~~Red~~ Medium and darker shades of grey ~~represents~~ represent expression greater than the median, black is equal to the median and ~~green~~

lighter shades of grey represent expression that is less than the median. Fig. 12B shows HOX genes median expression across the VxInsight clusters of the infant leukemia data set. The ~~red, blue~~ light grey, dark grey and black bars represent the median of expression of each HOX family gene across all the cases in VxInsight clusters A, B and C, respectively.

On page 11 of the BRIEF DESCRIPTION OF THE DRAWINGS amend the description of Figure 15 as follows:

Figure 15 shows genes that characterize the t(4;11) translocation in A vs. B, derived from the *VxInsight* clustering program using ANOVA. The ~~red color~~ shaded areas ~~represents~~ represent genes that have higher expression in the t(4;11) cases in *VxInsight* cluster A against the t(4;11) cases in *VxInsight* cluster B.

On page 11 of the BRIEF DESCRIPTION OF THE DRAWINGS amend the description of Figure 16 as follows:

Figure 16 shows genes that characterize each one of the *MLL* translocations (derived from Bayesian Networks Analysis). The ~~highlighted~~ shaded genes represent possible therapeutic targets.

On page 11 of the BRIEF DESCRIPTION OF THE DRAWINGS amend the description of Figure 18 as follows:

Figure 18 shows genes that characterize the t(4;11) translocation (left column)

and the *MLL* translocations (right column), derived from the *VxInsight* clustering program using ANOVA. The ~~red color~~ shaded areas ~~represents~~ represent genes that have higher expression in the t(4;11) cases against the rest of the cases or the MLL cases against the rest.

On page 173, in the DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS, amend the second full paragraph as follows:

To explore potential clusters driven by gene expression profiles, the initial analysis of the pediatric ALL cohort was accomplished using a force directed clustering algorithm coupled with a novel visualization tool, *VxInsight* as described in Example IB. Unexpectedly, we discovered 9 novel biologic clusters of ALL (2 distinct T-cell ALL clusters (Sl and S2) and 7 (2 related clusters are seen in cluster X) distinct B-lineage ALL clusters (A, B, C, X, Y, Z)) each with distinguishing gene expression profiles. (Fig. 6A) Using ANOVA, we identified over 100 statistically significant genes uniquely distinguishing each of these cohorts; a list of the top statistically significant genes distinguishing each cluster is provided in Table 43. Review of these lists of genes reveals many interesting signaling molecules and transcription factors. The X cluster (which contains two highly related clusters) is quite unique in having expression of several genes regulating methylation and folate metabolism.

On pages 173 and 174, in the DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS, amend the last paragraph bridging pages 173 and 174 as follows:

Examination of the cluster data reveals that while there are some trends, no cytogenetic abnormality precisely defines or is correlated with any specific cluster. (Fig. 6B) It is interesting that cases with a t(12;21) or hyperdiploidy, both conferring low risk and good outcomes, tend to cluster together; although combinations of these cases can be seen primarily in clusters C and Z as well as the top component of the X cluster indicating that there is still heterogeneity in gene expression profiles associated with these clusters. On the terrain map from VxInsight (~~Fig. 6~~ Fig. 6A, top) these three cluster regions (C, Z, and X) are actually fairly closely approximated indicating they are more related than for example cluster C to cluster S2. Although our correlations between outcome and clusters are still underway, it is interesting that the hyperdiploid and t(12;21) cases in cluster X had a significantly poorer outcome than those in cluster C or Z, suggesting that these cluster groupings may reflect different biologic propensities that confer differing responses to therapy. Similarly, the t(1;19) cases clustered in Y had a poorer outcome than those in clusters A and B. (Fig. 6B) Finally, it is of interest that ALL cases with t(9;22) simply don't cluster, they appear to be distributed among virtually all B precursor clusters. While we do not understand the significance of this result, it suggests that the t(9;22) is a pre-leukemic or initiating genetic lesion that may not be sufficient for leukemogenesis, or alternatively, that clones with a t(9;22) are quite genetically unstable and transformation and genetic progression may occur along many pathways. Results similar to our own were recently reported by Fine et al. (Blood Abstract, Blood Supplement 2002 (753a, Abstract #2979)). Using hierarchical clustering on a small series of 35 cell lines and ALL cases, these investigators found a limited

correlation between intrinsic biologic clusters in ALL and cytogenetic abnormalities; cases with a t(9;22) were found to be particularly heterogeneous in their gene expression profiles.

On pages 200 and 201, in the DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS, amend the last paragraph bridging pages 200 and 201 as follows:

The next technique used was Principal Component Analysis (PCA). PCA, closely related to the Singular Value Decomposition (SVD), is an unsupervised data analysis method whereby the most variance is captured in the least number of coordinates (Joliffe, 1986; Kirby, 2001; Trefethan & Bau, 1997). As shown in Fig. 9, the first three principal components can be seen to partition the infant cohort into two different groups (represented by two different shades of grey). These groups capture the infant ALL/AML lineage distinction, but only weakly agree with the MLL cytogenetics. Specifically, there is a 92% agreement between the PCA and the ALL/AML labels and only a 65% agreement between the PCA and MLL/non-MLL labels. Unexpectedly, the ALL/AML distinction does not appear until the second principal component, suggesting that morphology is not the most important factor explaining the variance in our data set. However, the first (and most important) principal component does not reveal any obvious clusters. Upon further analysis with a force-directed graph layout algorithm, we found ~~the~~ an additional group (discussed later) seen only in the first principal component (~~colored in blue~~ represented in solid black in the bottom row in Fig. 9).

On page 205, in the DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS, amend the first full paragraph as follows:

Finally, the third ~~rightmost~~ cluster (~~Fig. 9~~ Fig. 10 , cluster C, n=54, 42 AML cases and 12 ALL cases) is more heterogeneous and has a broader spectrum of *MLL* translocations. The gene expression signature of this group seems to have "myeloid" characteristics, with activation of genes previously reported as "myeloid-specific" such as Cystatin C (CST3), the myeloid cell nuclear differentiation factor (MNDA), and CCAAT/enhancer binding protein delta (C/EBP) (Golub, 1999; Skalnik, 2002). Members of the CCAAT/enhancer binding protein (C/EBP) family of transcription factors are important regulators of myeloid cell development (Skalnik, 2002). Other genes useful for cluster C prediction may also provide new insights into infant leukemia pathogenesis. For example, the mitogen activated protein kinase-activated protein kinase 3 is the first kinase to be activated through all 3 MAPK cascades: extracellular signal-regulated kinase (ERK), MAPKAP kinase-2, and Jun-N-terminal kinases/stress-activated protein kinases (Ludwig, 1996). It has been demonstrated as a determinant integrative element of signaling in both mitogen and stress responses. MAPKAPK3 showed high relative expression in the patients in cluster C. Many of the genes that characterize this cluster encode proteins characteristic of definitive myeloid differentiation (NDUFAB1, SOD1, GSTTLp28), or which are critical for signal

transduction (TYROBP). Interestingly, activation of many DNA repair and GST genes was also evident in this group of cases.

On pages 205 and 206, in the DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS, amend the last paragraph bridging pages 205 and 206 as follows:

The most common mutations in infant leukemia are translocations of the *MLL* gene at chromosome band 11q23. Interestingly, the *MLL* cases in cluster A (~~Fig. 10, lower left panel~~ Fig. 10C) are primarily t(4;11) (n=7), as well as two cases with t(10;11) and one with t(11;19). Cluster B, composed of virtually entirely ALL cases, contains a large number of t(4; 11) cases (n=29) as well as four cases with t(11;19), one case of t(10;11), and one case of t(1;11). Finally, ~~the bottom right~~ cluster C (n=54), predominantly AML but containing twelve cases with an ALL label that nonetheless have more "myeloid" patterns of gene expression, also comprises five cases with t(9;11), three cases with t(1;11), three cases with t(11;19), one case with t(4;11) and three cases with other MLL translocations.